



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 1 026 242 A1**

(12) **EUROPEAN PATENT APPLICATION**  
published in accordance with Art. 158(3) EPC

(43) Date of publication:  
**09.08.2000 Bulletin 2000/32**

(51) Int. Cl.<sup>7</sup>: **C12N 15/10**

(21) Application number: **98950331.3**

(86) International application number:  
**PCT/JP98/04772**

(22) Date of filing: **21.10.1998**

(87) International publication number:  
**WO 99/20750 (29.04.1999 Gazette 1999/17)**

(84) Designated Contracting States:  
**DE FR GB**

(30) Priority: **22.10.1997 JP 28998297**

(71) Applicant:  
**Helix Research Institute  
Chiba 292-0812 (JP)**

(72) Inventors:  
• **OTA, Toshio**  
**Kisarazu-shi Chiba 292-0801 (JP)**

- **NISHIKAWA, Tetsuo**  
**Kisarazu-shi Chiba 292-0833 (JP)**
- **SALAMOV, Asaf**  
**Saffron Walden Essex CB10 2AP (GB)**
- **ISOGAI, Takao**  
**Kisarazu-shi Chiba 292-0833 (JP)**

(74) Representative:  
**VOSSIUS & PARTNER  
Siebertstrasse 4  
81675 München (DE)**

(54) **METHOD FOR SCREENING FULL-LENGTH cDNA CLONES**

(57) A method for efficiently screening full-length cDNA clones which comprises: determining the base sequence in the 5'-region of each clone contained in a cDNA library prepared by a method for constructing a cDNA library involving full-length ones at a high ratio; examining the presence/absence of initiation ATG in this 5'-region and the location thereof by using an originally developed software for anticipating initiation codons in cDNA; thus exactly judging the presence/absence of the initiation codon and the location thereof; and screening the cDNAs thus judged as carrying the initiation codon from the cDNA library. Moreover, a cDNA library containing full-length ones at an extremely high ratio can be constructed by mixing the clones thus selected above.

**EP 1 026 242 A1**

**Description**Technical field

5 **[0001]** The present invention belongs to the field of genetic engineering, and relates to a method for screening full-length cDNA clones.

Background Art

10 **[0002]** Recently, genome projects targeting various animals, plants, and microorganisms have been in progress. Numerous genes have been isolated and their functions are under investigation. In order to efficiently analyze the functions of isolated genes, it is important to efficiently obtain cDNA clones capable of expressing complete proteins, that is, full-length cDNA clones.

**[0003]** The followings are known as methods for constructing a full length-enriched cDNA library: the oligo capping method in which an RNA linker is enzymatically bound to Cap of mRNA (Sugano & Maruyama, *Proteins, Nucleic Acids and Enzymes*, 38: 476-481, 1993, Suzuki & Sugano, *Proteins, Nucleic Acids and Enzymes*, 41: 603-607, 1996, M. Maruyama and S. Sugano, *Gene*, 138, 171-174, 1994); the modified oligo capping method developed by combining the oligo capping method with Okayama-Berg method (S. Kato et al., *Gene*, 150, 243-250, 1994, Kato & Sekine, *Unexamined Published Japanese Patent Application (JP-A) NO. Hei 6-153953*, published June 3, 1994); and the linker chemical-binding method in which a DNA linker is bound to Cap (N. Merenkova and D. M. Edwards, *WO 96/34981 Nov. 7, 1996*), the cap chemical modification method by biotin modification of Cap (P. Carninci et al., *Genomics*, 37, 327-336, 1996, P. Carninci et al., *DNA Research*, 4, 61-66, 1997). These are all methods to modify Cap of eukaryotic mRNA and to prepare a full length-enriched cDNA library. A known method for constructing a full length-enriched cDNA library by trapping Cap is the method using Cap-binding proteins derived from yeast or Hela cells for labeling a 5'-cap site (I. Edery et al., *MCB*, 15, 3363-3371, 1995). Also known is Cap Finder (Clontech) that is the Cap Switch oligonucleotide method in which the Cap Switch oligonucleotide is annealed by C-tailing the 5' end of a first strand cDNA.

25 **[0004]** A cDNA library constructed by these methods is rich in full-length cDNAs compared to that obtained by the conventional methods. However, incomplete-length clones are also contained to some extent. To efficiently analyze the functions of genes and to efficiently clone novel useful genes, development of methods for easily confirming whether each clone contained in a cDNA library is full-length or not has been desired.

Disclosure of the Invention

35 **[0005]** An objective of the present invention is to provide a method for efficiently screening full-length cDNA clones, and a method for constructing a full length-enriched cDNA library.

**[0006]** The present inventors have studied to achieve the above objective and contemplated efficiently screening full-length cDNAs from a cDNA library by the presence or absence of a translation initiation codon as an index based on the fact that a cDNA deficient in a certain 5'-region is likely to lack a translation initiation codon, whereas a full-length cDNA contains an initiation codon. Specifically, the inventors assumed that a full-length cDNA could be efficiently screened from a cDNA library constructed by a method for preparing a full length-enriched cDNA library. Specifically, the inventors thought that full-length cDNA clones could be efficiently isolated by constructing a cDNA library by a method for preparing a full length-enriched cDNA library, determining several hundreds of base pairs of a DNA nucleotide sequence from the 5' end, and analyzing the presence or absence of an initiation codon in this region to screen the clones containing initiation codons.

45 **[0007]** However, few programs for predicting an initiation site of cDNA have been developed (e.g., "A. G. Pedersen, *Proceedings of fifth international conference on intelligent systems for molecular biology*, p226-233, 1997, held in Halkidiki, Greece, June 21-26, 1997). Though some programs for exons prediction have been developed ("Gene Finder". V. V. Solovyev et al., *Nucleic Acids Res.*, 22, 5156-5163, 1994, "Grail" Y. Xu et al., *Genet-Eng-N-Y.*, 16, 241-253, 1994), an initiation site cannot be accurately determined relying solely on these programs.

50 **[0008]** The present inventors have developed a program for cDNA initiation codon prediction by themselves and determined nucleotide sequences of the 5'-region of clones contained in a cDNA library constructed by a method for preparing a full length-enriched cDNA library to examine whether an initiation codon exists in this 5'-region using this software program.

**[0009]** More specifically, a full length-enriched cDNA library was constructed by the oligo capping method and nucleotide sequences of the 5'-regions of some clones contained in the cDNA library were determined. Based on the determined sequences, the clones were divided into known and novel ones through a database search. The presence or absence of an initiation codon and its location in the determined nucleotide sequences of the 5'-regions were judged using the initiation codon prediction program. For the known clones, whether the location of the initiation codon recog-

nized by the initiation codon prediction program coincides with that of the initiation codon in databases is examined. Indeed, the presence or absence and location of the initiation codon in the known clones predicted by the program coincided with the information in the databases.

**[0010]** Thus, the software program developed by the present inventors can accurately recognize the presence or absence of an initiation codon and its location, and full-length cDNA clones can be efficiently screened by selecting the clones that are recognized to contain an initiation codon by the program from the cDNA library. Moreover, a cDNA library extremely rich in full-length cDNAs can be constructed by combining the screened clones.

**[0011]** The present invention relates to a method for screening full-length cDNA clones from a cDNA library and a method for constructing a full-length cDNA library by combining cDNA clones screened by the screening method. More specifically, it relates to:

(1) A method for isolating a full-length cDNA clone, the method comprising:

- (a) determining a nucleotide sequence from the 5'-region of a cDNA clone contained in a cDNA library,
- (b) determining the presence or absence of an initiation codon in the nucleotide sequence determined in (a) using an initiation codon prediction program, and
- (c) selecting clones recognized as containing the initiation codon in (b);

(2) The method of (1), wherein the cDNA library is constructed by a method for preparing a full length-enriched cDNA library;

(3) The method of (1), wherein a cDNA library is constructed by a method comprising a step of modifying Cap of mRNA;

(4) A method for constructing a full length cDNA library, the method comprising:

- (a) determining a nucleotide sequence from the 5'-region of a cDNA clone contained in a cDNA library,
- (b) determining the presence or absence of an initiation codon in the nucleotide sequence determined in (a) using an initiation codon prediction program,
- (c) selecting clones recognized as containing the initiation codon in (b), and
- (d) combining the clones selected in (c);

(5) The method of (4), wherein the cDNA library is prepared by a method for constructing a full length-enriched cDNA library;

(6) The method of (4), wherein the cDNA library is constructed by a method comprising a step of modifying Cap of mRNA; and

(7) A cDNA library obtainable by the method of (4).

**[0012]** The present invention is based on the inventors' findings that full-length cDNA clones can be efficiently isolated by analyzing nucleotide sequences of the 5'-region of cDNAs in a cDNA library, specifically a full length-enriched cDNA library, by using a software program for accurately predicting a translation initiation codon, and a full length-enriched cDNA library can be constructed by combining the isolated cDNA clones. The method for screening full-length cDNA clones by the present invention comprises (a) determining a nucleotide sequence from the 5'-region of a cDNA clone contained in a cDNA library, (b) determining the presence or absence of an initiation codon in the determined nucleotide sequence using an initiation codon prediction program, and (c) selecting clones recognized as containing the initiation codon. The method for constructing a full-length cDNA library of the present invention comprises, in addition to above steps (a) to (c), step (d) of combining the screened clones.

**[0013]** In the method of the present invention, a "cDNA clone" whose nucleotide sequence of the 5'-region is to be determined is not particularly limited. Full-length cDNAs cannot be efficiently isolated from clones derived from a library not rich in full-length cDNAs, compared with clones derived from a full length-enriched cDNA library. Therefore, a cDNA clone is preferably derived from a library constructed by the above-described methods for preparing a full length-enriched cDNA library, including, for example, the oligo capping method in which an RNA linker is enzymatically bound to Cap of mRNA (Sugano & Maruyama, *Proteins, Nucleic Acids and Enzymes*, 38: 476-481, 1993, Suzuki & Sugano, *Proteins, Nucleic Acids and Enzymes*, 41: 603-607, 1996, M. Maruyama and S. Sugano, *Gene*, 138, 171-174, 1994), the modified oligo capping method developed by combining the oligo capping method with Okayama-Berg method (S. Kato et al, *Gene*, 150, 243-250, 1994, Kato & Sekine, JP-A-Hei 6-153953, June 3, 1994), the linker chemical-binding method in which a DNA linker is chemically bound to Cap (N. Merenkova and D. M. Edwards, WO 96/34981 Nov. 7, 1996), the Cap chemical modification method in which Cap is modified with biotin (P. Carninci et al., *Genomics*, 37, 327-336, 1996, P. Carninci et al., *DNA Research*, 4, 61-66, 1997), the method using Cap binding proteins derived from yeast or HeLa cells (I. Ederly et al, *MCB*, 15, 3363-3371, 1995), or a library prepared by Cap Finder using Cap Switch oligo-

nucleotide method.

[0014] A cDNA clone can be isolated from a cDNA library by standard methods described in, for example, J. Sambrook, E. F. Fritsch & T. Maniatis, Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press, 1989.

[0015] A nucleotide sequence can be determined from the 5'-region of a clone by, for example, standard methods using DNA sequencing reagents and a DNA sequencer available from Applied Biosystems, etc. A whole nucleotide sequence of the clone does not have to be determined, and determining about 1,000 nucleotides from the 5' end is sufficient. The high accuracy can be expected by determining about 500 nucleotides, even about 300 nucleotides.

[0016] An "initiation codon prediction program" used for analyzing a nucleotide sequence from the 5'-region of a clone is preferably the program developed by the present inventors as described in Example 1 below. The presence or absence of an initiation codon in a determined sequence is judged by a score deduced from the results of analysis with the program. A cDNA clone with a high score, recognized as containing an initiation codon in the determined sequence, is usually comprised of a full-length cDNA, while one with a low score, recognized as not containing an initiation codon in the determined sequence, contains an incomplete-length cDNA. Thus, a full-length cDNA can be efficiently isolated by screening a cDNA from a cDNA library, judged as containing an initiation codon in the nucleotide sequence. Indeed, in one embodiment of the analysis with the program described in Example 1 below where a cDNA library with the full-length cDNA content of 51% was used to screen clones (the highest score was 0.94), the content of full-length clones among the screened clones was 71% when clones showing a score of 0.5 or higher were selected, 77% with a score of 0.70 or higher, 81% with a score of 0.80 or higher, and 85% with a score of 0.90 or higher. Therefore, full-length cDNA clones can be screened with a high accuracy by selecting clones with high scores using the program described in Example 1.

[0017] Moreover, a cDNA library re-constructed by combining clones selected by the method for screening full-length cDNA clones of the present invention is extremely rich in full-length cDNAs compared with the parent cDNA library used for screening clones. By expressing whole cDNAs capable of expressing proteins in the thus-obtained library, a system for efficiently analyzing gene functions containing a mixture of expressed proteins can be obtained. This system enables efficiently cloning useful genes.

#### Best Mode for Carrying out the Invention

[0018] The present invention is illustrated in detail below with reference to the following examples, but is not to be construed as being limited thereto.

#### Example 1. Preparation of a program for predicting a translation initiation codon of cDNA

[0019] The translation initiation codon prediction program of the present invention recognizes a putative authentic initiation codon among all ATGs contained in a given cDNA sequence fragment. The program predicts based on A) information on similarity of given regions (several tens to several hundreds base pairs) at both sides of a putative ATG to translational regions and B) information on similarity of regions near a putative ATG to those near an authentic initiation codon. Characteristics of sequences in a translational region and regions near an initiation codon are extracted beforehand by from information of numerous sequences whose translational and non-translational regions have been identified. The program predicts an initiation codon based on the information about the above characteristics.

[0020] The linear discriminant analysis used in Gene Finder, a program for genomic exon prediction (Solovyev V. V., Salamov A. A., Lawrence C. B. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic. Acids Res, 1994, 22: 5156-63), was applied to optimize prediction. In the linear discriminant analysis, information on some characteristics derived from data is digitized, weighted, and then calculated a score. Here, a score is converted into a probability of similarity to an initiation codon (the probability is a rate of correct answers obtained from data of sequences whose initiation codon has been identified). Specifically, a probability of similarity to an initiation codon of each ATG contained in a given cDNA sequence is output. Recognition as an initiation codon is determined whether a probability of similarity to an initiation codon is above a given threshold value or not. A threshold value is established depending on the plan of the following analyses, that is, depending on the extent of noises acceptable for the following analysis. For example, when 40% of noise is acceptable, a threshold value of 0.6 can be used. A parameter of weight is determined so as to maximize the prediction system using data of sequences whose initiation codon has been identified as a training datum. The above information of A) and B) were each embodied into the following three information and used as information about characteristics.

A) information on similarity of given regions (several tens to several hundreds base pairs) at both sides of a putative ATG to translational regions

[0021]

5

1: a frequency of six nucleotide base letters contained in a sequence from ATG to a stop codon (within 300 bp downstream of ATG at longest)

2: discrepancy of the information on a frequency of six nucleotide base letters contained in 50 nucleotide bases upstream and downstream of ATG

10

3: an index of similarity to a signal peptide [a hydrophobicity index of the most hydrophobic eight amino acids letters among 30 amino acids (90 nucleotide bases) downstream of ATG]

B) information on similarity of regions near a putative ATG to those near an authentic initiation codon

15 [0022]

1: information on a weighted matrix as using three nucleotide base letters in the region from 14 nucleotide bases upstream of ATG to 5 nucleotide bases downstream of ATG as a unit

20

2) the presence or absence of other ATGs upstream of ATG in a same frame (the presence is 1 and the absence is 0)

3: a frequency of cytosine contained in the region from 36 bases upstream of ATG to 7 bases downstream of ATG.

Example 2: Preparation of cDNA by the oligo capping method and analysis thereof by the program for initiation codon prediction

25

[0023] A cDNA library was prepared by the oligo capping method and the plasmid DNA was extracted from each clone by the standard method. Specifically, mRNA was extracted from human placenta and human cultured cells (Tetradocarcinoma NT-2 and neuroblastoma SK-N-MC) by the method described in the reference (J. Sambrook, E. F., Fritsch & T. Maniatis, Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press, 1989). An oligo cap linker (SEQ ID NO. 1) with an oligo dT adaptor primer (SEQ ID NO. 2) in the case of Tables 1 & 2, or with a random adaptor primer (SEQ ID NO. 3) in the case of Tables 3 & 4 were subjected to BAP treatment, TAP treatment, RNA ligation, synthesis of a first strand cDNA, and removal of RNA according to the methods described in the references (Suzuki & Sugano, Proteins, Nucleic Acids, and Enzymes, 41, 603-607, 1996, p606, Y. Suzuki et al., Gene, 200, 149-156, 1997). The first strand cDNA was then converted into the double-stranded DNA by PCR, digested with *Sfi*I, and cloned into vectors, such as pME18SCG, pMFL etc. digested with *Dra*III in the determined direction (Sugano & Maruyama, Proteins, Nucleic Acids, and Enzymes, 38, 472-481, 1993, p480). The obtained DNA was subjected to the sequencing reaction using a DNA sequencing reagent (DyeTerminator Cycle Sequencing FS Ready Reaction Kit, PE Applied Biosystems) following the manual and sequenced with a DNA sequencer (ABI PRISM 377, PE Applied Biosystems). The DNA sequence of the 5'-region of each clone was analyzed once.

40

[0024] The presence or absence of an initiation codon in the DNA sequence of each clone was analyzed using the developed program for cDNA initiation codon prediction (ATGpr). In this analyzing program, the higher the score is, the higher the probability of being an initiation codon is. The maximum score is 0.94.

(1) Analysis of translation initiation codons in the clones whose open reading frames are known in database among cDNA prepared by the oligo capping method

45

[0025] Among the results for all analyzed clones, the result for the clones that are known to contain the initiation codon in the determined sequences in databases (F-NT2RP1000020, F-NT2RP1000025, F-NT2RP1000039, and F-NT2RP1000046) are shown in Table 1. F-NT2RP1000020 (880 bp) has 96% identity at nucleotide positions 88 to 690 to "human neuron-specific gamma-2 enolase" (GenBank accession No. M22349); F-NT2RP1000025 (645 bp), 97% homology at positions 29 to 641 to "human alpha-tubulin mRNA" (GenBank accession No. K00558); F-NT2RP1000039 (820 bp), 96% identity at positions 12 to 820 to "human mRNA for elongation factor 1 alpha subunit (EF-1 alpha)" (GenBank accession No. X03558); and F-NT2RP1000046 (788 bp), 97% identity at positions 3-788 to "human M2-type pyruvate kinase mRNA" (GenBank accession No. M23725). The sequences of the 5'-region in these clones are shown in SEQ ID Nos: 4, 5, 6, and 7.

55

Table 1

ATG No.	F-NT2RP1000020		F-NT2RP1000025		F-NT2RP1000039		F-NT2RP1000046	
	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
1	1	0.05	96	(0.94)	65	(0.90)	111	(0.94)
2	162	(0.84)	148	0.13	154	0.05	174	0.82
3	292	0.05	193	0.05	209	0.11	198	0.19
4	313	0.05	201	0.09	231	0.05	300	0.16
5	441	0.05	232	0.05	321	0.05	315	0.11
Note 1: ( ) means translation initiation codon Note 2: Location of ATG means the nucleotide base position of ATG in the 5'-region of a DNA sequence. ATG No. means the number of ATG from the 5'-region of a DNA sequence.								

As show in Table 1, among the cDNA prepared by the oligo capping method, the full-length clones whose open reading frames are known in databases, containing initiation codons were accurately recognized by the initiation codon prediction program (ATGpr) (coincident with the initiation codons in databases).

(2) Analysis of initiation codons in the clones whose open reading frames are known in database among cDNA prepared by the oligo capping method

Among the results for the clones analyzed, the results for the clones whose initiation codon is known to absent in the determined sequence in databases (F-NT2RP1000013, F-NT2RP1000054, and F-NT2RP1000122) are shown in Table 2. F-NT2RP1000013 (608 bp) has 97% identity at positions 1 to 606 to "human nuclear matrix protein 55 (nmt55) mRNA" (GenBank accession No. U89867); F-NT2RP1000054 (869 bp), 96% identity at positions 1 to 869 to "human signal recognition particle (SRP54) mRNA" (GenBank accession No. U51920); and F-NT2RP1000122 (813 bp), 98% identity at positions 1 to 813 to "*H. sapiens* mRNA for 2-5A binding protein" (GenBank accession No. X76388). The sequences of the 5' region of these clones are shown in SEQ ID Nos: 8, 9, and 10.

Table 2

ATG No.	F-NT2RP1000013		FNT2RP1000054		F-NT2RP1000122	
	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
1	21	0.05	31	0.12	23	0.07
2	27	0.05	60	0.20	100	0.05
3	32	0.32	87	0.05	166	0.05
4	56	0.11	97	0.05	235	0.06
5	119	0.10	146	0.05	316	0.05
6	125	0.08	172	0.05	346	0.05
7	141	0.05	180	0.11	406	0.05
8	155	0.06	218	0.07	431	0.05
9	161	0.06	272	0.05	469	0.06
10	176	0.08	319	0.07	546	0.12
11	203	0.07	346	0.05	553	0.05
12	290	0.20	363	0.07	574	0.05

Table 2 (continued)

ATG No.	F-NT2RP1000013		FNT2RP1000054		F-NT2RP1000122	
	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
13	311	0.16	409	0.05		
14	314	0.12	480	0.07		

[0028] As shown in Table 2, among cDNA prepared by oligo capping method, the initiation codon prediction program (ATGpr) did not recognize by mistake the initiation codons in incomplete-length cDNAs whose open reading frames are known in databases and which do not contain any initiation codons.

(3) Analysis of initiation codons in novel clones among the cDNA prepared by the oligo capping method

[0029] Among the results for analyzed clones, the results for novel clones that were predicted to contain initiation codons (F-ZRV6C1000408, F-ZRV6C1000454, F-ZRV6C1000466, F-ZRV6C1000615, and F-ZRV6C1000670) are shown in Table 3. The sequences of the 5' region of these clones are shown in SEQ ID Nos: 11, 12, 13, 14, 15.

Table 3

F-ZRV6C1000408			F-ZRV6C1000454		F-ZRV6C1000466	
ATG No.	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
1	85	<0.94>	5	0.05	162	<0.86>
2	208	0.22	107	<0.87>	182	0.05
3	386	0.05	153	0.05	207	0.08
4	518	0.11	201	0.08	244	0.05
5	545	0.05	211	0.05	262	0.05
6			236	0.07	303	0.11

(cont'd)

Table 3 (cont'd)

F-ZRV6C1000615			F-ZRV6C1000670	
ATG No.	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
1	85	<0.94>	120	<0.94>
2	208	0.26	187	0.54
3	386	0.05	312	0.06
4	518	0.09	388	0.05
5	545	0.05	445	0.05

Note: <> means predicted initiation codon.

[0030] As shown in Table 3, the predicted initiation codons in F-ZRV6C1000408, F-ZRV6C1000454, F-ZRV6C1000466, F-ZRV6C1000615, and F-ZRV6C1000670 are "ATG" starting with "A" at positions 85, 107, 162, 85, and 120, respectively. Therefore, these clones were judged as full-length cDNA clones.

# EP 1 026 242 A1

[0031] In addition, among the results for analyzed clones the results for novel clones predicted as not containing initiation codons (F-ZRV6C1001410, F-ZRV6C1001197, and F-ZRV6C1001472) are shown in Table 4. The sequences of the 5' region of these clones are shown in SEQ ID Nos: 16, 17 and 18.

Table 4

ATG No.	F-ZRV6C1001410		F-ZRV6C1001197		F-ZRV6C1001472	
	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score	Location of ATG	ATGpr Score
1	23	0.05	5	0.24	77	0.25
2	31	0.07	141	0.25	126	0.05
3	71	0.06	202	0.05	149	0.05
4	178	0.05	219	0.05	194	0.05
5	214	0.05	228	0.05	213	0.22
6					249	0.05
7					338	0.09
8					344	0.05
9					351	0.05
10					365	0.05

[0032] As shown in Table 4, F-ZRV6C1001410, F-ZRV6C1001197, and F-ZRV6C1001472 were recognized as not containing initiation codons. These clones were thus judged as incomplete-length clones.

## Industrial Applicability

[0033] The present invention provides a method for efficiently selecting full-length cDNAs. Clones selected by the method of the present invention can express complete proteins. Therefore, the present invention enables efficiently analyzing the functions of isolated genes.



SEQUENCE LISTING

5

<110> Helix Research Institute, Inc.

10

<120> Method for screening full-length cDNA clones

<130> H1-806PCT

15

<150> JP 09-289982

<151> 1997-10-22

<160> 18

20

<170> PatentIn version 2.0

25

<210> 1

<211> 30

<212> DNA

<213> Artificial Sequence

30

<220>

<223> Oligo-capping linker sequence

35

<400> 1

AGCAUCGAGU CGGCCUUGUU GCGCUACUGG

30

40

<210> 2

<211> 42

<212> DNA

<213> Artificial Sequence

45

<220>

<223> Oligo(dT) adapter primer sequence

50

<400> 2

GCGGCTGAAG ACGGCCTATG TGGCCTTTT TTTTTTTT TT

42

55

<210> 3

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Random adapter primer sequence

<400> 3

GCGGCTGAAG ACGGCCTATG TGGCCNNNN NC

32

<210> 4

<211> 880

<212> DNA

<213> Homo sapiens

<400> 4

ATGCGCCCGC GCGGCCCTAT AGGCGCCTCC TCGCCCGGCC GCCCGGGAGC CGCAGCCGCC 60  
 GCGGCCCACTG CCACTCCCGC TCTCTCAGCG CCGCCGTGCG CACCGCCACC GCCACTGCCA 120  
 CTACCACCGT CTGAGTCTGC AGTCCCGAGA TCCCAGCCAT CATGTCCATA GAGAAGATCT 180  
 GGGCCCGGGA GATCCTGGAC TCCCGCGGGA ACCCCACAGT GGAGGTGGAT CTCTATACTG 240  
 CCAAAGGTCC TTTCCGGGCT GCAGTGCCCA GTGGAGCCTC TACGGGCATC TATGAGGCCC 300  
 TGGAGCTGAG GGATGGAGAC AAACAGCGTT ACTTAGGCAA AGGTGTCCTG AAGGCAGTGG 360  
 ACCACATCAA CTCCACCATC GCGCCAGCCC TCATCAGCTC AGGTCTCTCT GTGGTGGAGC 420  
 AAGAGAAACT GGACAACCTG ATGCTGGAGT TGGATGGGAC TGAGAACAAA TCCAAGTTTG 480  
 GGGCCAATCC ATCCTGGGTG TGTCTCTGGC CGTGTGTAAG GCANGGGCAA CTGAACNGGA 540  
 ACTGCCCTG TATCGCCACA TTGCTCAGCT TGGNCGGGA CTCANACCTC ATCCTGCCTG 600  
 TTGCCGGCCT TCAACGTGAT CAATGGTTGG CTTCTCATGC CTGGCAACAA ANCTGGCCAT 660  
 TGCNGGAATT TTCATGATCC TCCCCNTTGG GAAACTGAAA AACTTTCCGG AATGCCCNTC 720  
 CAACTAAGTT GCAAAAGGTC TACCNATACC CCCCAAGGGG AATTCCTCCA AGGGAACAAA 780  
 TNCCCGGGAA AGGAATGCCC CCCAATTNTT NCGGGGAATA AAAGGTGGGC TTTGCCCCCC 840  
 CATTTTCCTG GAAAAACNA TNAACCCT TGGGAACTT 880

<210> 5

&lt;211&gt; 645

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

10 TGTGCGTTAC TTACCTCNAC TCTTAGCTTG TCGGGGACGG TAACCGGGAC CCGGTGTCTG 60  
 CTCCTGTGCG CTTGCGCTCC TAATCCCTAG CCACTATGCG TGAGTGCATC TCCATCCACG 120  
 TTGGCCAGGC TGGTGTCCAN ATTGGCAATG CCTGCTGGGA GCTCTACTGC CTGGAACACG 180  
 15 GCATCCAGCC CGATGGCCAG ATGCCAAGTG ACAAGACCAT TGGGGGAGGA GATGACTCCT 240  
 TCAACACCTT CTTCACTGAG ACGGGCGCTG GCAANCACGT GCCCGGGGCT GTGTTTGTAG 300  
 ACTTGAACC CACAGTCATT GATGAAGTTC GCACTGGCAC CTACCGCCAG CTCTCCACC 360  
 CTGAGCAGCT CATCNCAGGC AAGGAAGATG CTGCCAATAA CTATGCCCGA GGGCACTACA 420  
 20 CCATTGGCAA GGAGATCATT GACCTTGTGT TGGACCGAAT TCGCAAGCTG GCTGACCANT 480  
 GCACCGGTCT TCANGGCTTC TTGGTTTCC ACAGCTTTGG TGGGGGAAGT GGTTCGGGT 540  
 TCACCTCCCT GCTCATGGAA CGTCTCTCAG TTGATTATGG CAAGAAATCC AAGCTGGAGT 600  
 25 TCTCCATTTA CCCAGCACCC CNGGTTTCN CNGCTGTANT TNGAA 645

&lt;210&gt; 6

&lt;211&gt; 820

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 6

CTTTTTTCGC AACGGGTTTG CCGCCAGAAC ACAGGTGTCG TGAAAACACTAC CCCTAAAAGC 60  
 CAAAATGGGA AAGGAAAAGA CTCATATCAA CATTGTCGTC ATTGGACACG TAGATTCGGG 120  
 40 CAAGTCCACC ACTACTGGCC ATCTGATCTA TAAATGCGGT GGCATCGACA AAAGAACCAT 180  
 TGAAAAATTT GAGAAGGAGG CTGCTGAGAT GGGAAAGGGC TCCTTCAAGT ATGCCTGGGT 240  
 CTTGGATAAA CTGAAAGCTG AGCGTGAACG TGGTATCACC ATTGATATCT CCTGTGGAA 300  
 45 ATTTGAGACC AGCAAGTACT ATGTGACTAT CATTGATGCC CCAGGACACA GAGACTTTAT 360  
 CAAAAACATG ATTACAGGGA CATCTCAGGC TGAAGTGTCT GTCCTGATTG TTGCTGCTGG 420  
 TGTGTGGTAA TTTGAAGCTG GTATCTCCAA GAATGGGCAG ACCCGAGAGC ATGCCCTTCT 480  
 GGCTTACACA CTGGGTGTGA AACAACTAAT TGTGGGTGTT AACAAAATGG ATTCACTGAN 540  
 50 CCACCTTACA GCCAGAAGAA ATATGANGAA ATTGTAAAG AAGTCAGCAC TTACATTAAG 600  
 AAAATTGGCT ACAACCCCGA CACAGTANCA TTTGTGCCAA TTTCTGGTTG GAATGGTGAC 660

AACATGCTGG AACCAANTGC TAACATGCCT TGGTTCCAGG GATGGAAAAT CCCCNTTAA 720  
 GGATGGCNAT GCCATTGGAA CCCCCTGCT TGAAGGCTCT GGANTGCATC CTANCACCAA 780  
 CTCCTTCAAA TTGAAAAACC CCTTGCNCCC GCTCCNCCA 840

<210> 7

<211> 788

<212> DNA

<213> Homo sapiens

<400> 7

GAGGCTGAGG CAGTGGCTCC TTGCACAGCA GCTGCACGCG CCGTGGCTCC GGATCTCTTC 60  
 GTCTTTGCAG CGTAGCCCGA GTCGGTCAGC GCCGGAGGAC CTCAGCAGCC ATGTCGAAGC 120  
 CCCATAGTGA AGCCGGGACT GCCTTCATTC AGACCCAGCA GCTGCACGCA GCCATGGCTG 180  
 ACACATTCCT GGAGCACATG TGCCGCCTGG ACATTGATTC ACCACCCATC ACAGCCCGGA 240  
 ACACTGGCAT CATCTGTACC ATTGGCCCAG CTTCCCGATC AGTGGAGACG TTGAAGGAGA 300  
 TGATTAAGTC TGAATGAAT GTGGCTCGTC TGAACCTCTC TCATGGAACT CATGAGTACC 360  
 ATGCGGAGAC CATCAAGAAT GTGCGCACAG CCACGGAAAG CTTTGCTTCT GACCCCATCC 420  
 TCTACCGGCC CGTTGCTGTG GCTCTAGACA CTAAGGACC TGAGATCCGA ACTGGGCTCA 480  
 TCAAGGCGAG CGGCACTGCA GAGGTGGAGC TGAAGAATGG AGCCACTCTC AAAATCACGC 540  
 TGGATAATGC CTACATGGAA AAGTGTGACG AGAACATCCT GTGGCTGGAC TACAAGAACA 600  
 TCTGCAAGGT GGTGGAAGTG GGCAACAAGA TCTACGTGGA TGATGGGCTN ATTTCTCTCC 660  
 AGGTGAACAC AAAGGTGCCG ACTTCTGGG TGACNGANGT GGAATGGT GGCTCCTTGG 720  
 GCNCAAGAAA GGTGTGAACT TCCTGGGGCT GCTGTGGANT TGCCTGCTGT GTCNGAAAAA 780  
 GACATCCA 788

<210> 8

<211> 608

<212> DNA

<213> Homo sapiens

<400> 8

ACAGCCTGGC TCCTTTGAGT ATGAATATGC CATGCGCTGG AAGGCACTCA TTGAGATGGA 60  
 GAAGCAGCAG CAGGACCAAG TGGACCGCAA CATCNAGGAG GCTCGTGAGA AGCTGGAGAT 120  
 GGAGATGGAA GCTGCACGCC ATGAGCACCA GGTGATGCTA ATGAGACAGG ATTTGATGAG 180

5 GCGCCAAGAA GAACTTCGGA GGATGGAAGA GCTGCACAAC CAAGANGTGC AAAAACGAAA 240  
 GCAACTGGAG CTCAGGCAGG AGGAANAGCG CAGGCGCCGT GAAGANAGA TGCGGCGGCA 300  
 GCAAGAAGAA ATGATGCGGC GACNGCAGGA AGGATTCAAG GGAACCTTCC CTGATGCGAG 360  
 AGAGCAGGAG ATTCGGATGG GTCNGATGGC TATGGGAGGT GCTATGGGCA TAAACNACAG 420  
 ATGTGCCATG CCCCTGCTC CTGTGCCAGC TGGTACCCCA GCTCCTCCAG GACCTGCCAC 480  
 10 TATTATCCCG GATGGAAC TTGGGATTGAC CCCACCNACA ACTGAACGCT TTGGTCNGGC 540  
 TGCTACNATG GAANGAATTG GGGCAATTGG TGGAATCCT CCTGCATTCT ACCGTGCAGC 600  
 TCCTGGGA 608

15 <210> 9

<211> 869

20 <212> DNA

<213> Homo sapiens

25 <400> 9  
 ATATTAACT AGTGAAGCAA CTAAGAGAAA ATGTTAAGTC TGCTATTGAT CTTGAAGAGA 60  
 TGGCATCTGG TCTTAACAAA AGAAAAATGA TTCAGCATGC TGTATTTAAA GAACTTGTGA 120  
 AGCTTGTAGA CCCTGGAGTT AAGGCATGGA CACCCACTAA AGGAAAACAA AATGTGATTA 180  
 30 TGTTTGTGG ATTGCAAGGG AGTGGTAAAA CAACAACATG TTCAAAGCTA GCATATTATT 240  
 ACCAGAGGAA AGGTTGGAAG ACCTGTTTAA TATGTGCAGA CACATTCAGA GCAGGGGCTT 300  
 TTGACCAACT AAAACAGAAT GCTACCAAAG CAAGAATTCC ATTTTATGGA AGCTATACAG 360  
 AAATGGATCC TGTCAATCATT GCTTCTGAAG GAGTAGAGAA ATTTAAAAAT GAAAAATTTG 420  
 35 AAATTATTAT TGTGATACA AGTGGCCGCC ACAACAAGA AGACTCTTTG TTTGAAGAAA 480  
 TGCTTCAAGT TGCTAATGCT ATACAACCTG ATAACATTGT TTATGTGATG GATGCCTCCA 540  
 TTGGGCAGGC TTGTGAAGCC CAGGCTAAGG CTTTAAAGA TAAAGTAGAT GTACCTCAGT 600  
 AATAGTGACA AAACCTTGATG GCCATGCAAA ANGAAGTGGT GCACTCAGTG CAGTCGCTGC 660  
 40 CACAAAAAAT CCGATTATTT TCATTGGTAC AGGGGAACA TATANATGAC TTTGAACCTT 720  
 TCAAAAACAC AGCCTTTTAT TAACAAACTT CTTGGTATNG GCGACATTGA AAGGACTGAT 780  
 AAATAAAGTC CACNAATTGA AATTTGGATG ACNATGNAAA CCCTTATTGA AAAAATTGAA 840  
 45 ACATNGTCCA GTTTACTTT GCGAAACNT 869

50 <210> 10

<211> 813

<212> DNA

55

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

GTTGTGGTAT CTGTATTAAG AAATGCCCTT TTGGCGCCTT ATCAATTGTC AATCTACCAA 60  
 GCAACTTGGA AAAAGAAACC ACACATCGAT ATTGTGCCAA TGCCTTCAAA CTCACAGGT 120  
 TGCCTATCCC TCGTCCAGGT GAAGTTTGG GATTAGTTGG AACTAATGGT ATTGGAAAGT 180  
 CAACTGCTTT AAAAATTTTA GCAGGAAAAC AAAAGCCAAA CCTTGGAAAG TACGATGATC 240  
 CTCCTGACTG GCAGGAGATT TTGACTTATT TCCGTGGATC TGAATTACAA AATTACTTTA 300  
 CAAAGATTCT AGAAGATGAC CTAAGGCCA TCATCAAACC TCAATATGTA GACCAGATTC 360  
 CTAAGGCTGC AAAGGGGACA GTGGGATCTA TTTTGGACCG AAAAGATGAA ACAAAGACAC 420  
 AGGCAATTGT ATGTCAGCAG CTTGATTAA CCCACCTAAA AGAACGAAAT GTTGAAGATC 480  
 TTTCAGGAGG AGAGTTGCAG AGATTGCTT GTGCTGTCGT TTGCATACAG AAAGCTGATA 540  
 TTTTCATGTT TGATGAGCCT TCTAGTTACC TAGATGTCAA GCAGCGTTA AAGGCTGCTA 600  
 TTAATAACG ATCTCTAATA AATCCAGATA GATATATCAT TGTGGTGAA CATGATCTAA 660  
 GTGTATTAGA CTATCTCTCC GACTTCATCT GCTGTTTATA TGGTGATCA AGCGCCTATG 720  
 GAATTGTCAC TATGCCTTT AGTGTTAGAA AAGGCATAAA CNTTTTTTGG ATGGGTATGT 780  
 TCCAACAGAA AACTTGANAA TCNNAATGC NTC 813

&lt;210&gt; 11

&lt;211&gt; 655

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 11

AGACTCTCAC CGCAGCGGCC AGGAACGCCA GCCGTTACG CGTTCGGTCC TCCTTGGCTG 60  
 ACTCACCGCC CTCGCCGCG CACCATGGAC GCCCCAGGC AGGTGGTCAA CTTTGGGCCT 120  
 GGTCCCCGCA AGCTCCGCA CTCAGTGTG TTAGAGATAC AAAAGGAATT ATTAGACTAC 180  
 AAAGGAGTTG GCATTAGTGT TCTTGAAATG AGTCACAGGT CATCAGATT TGCCAAGATT 240  
 ATTAACAATA CAGAGAATCT TGTGCGGGAA TTGCTAGCTG TTCCAGACAA CTATAAGGTG 300  
 ATTTTCTGTC AAGGAGGTGG GTGCGGCCAG TTCAGTGCTG TCCCCTTAAA CCTCATTGGC 360  
 TTGAAGCAG GAAGGTGTGC GGAATATGTG GTGACAGGAG CTGGTCAGC TAAGGCCGCA 420  
 GAAGAAGCCA AGAAGTTTGG GACTATAAAT ATCGTTCACC CTAAACTTGG GAGTTATACA 480  
 AAAATTCCAG ATCCAAGCAC CTGGAACCTC AACCCANATG CCTCCTACGT GTTTTATTGC 540  
 NCAAAAGAAA CGGTGCATGG TGTGANTTT GACTTTATAC CCNATGTCAA GGAACANTA 600  
 CTGGTTTGTG ACATTTTCTT CCAACTTCTT GTCCAANCCA ATTGNATGTT TCCAA 655

<210> 12

5

<211> 599

<212> DNA

<213> Homo sapiens

10

<400> 12

15

20

25

```

AAAGATGCGC AGGCGCCGTG TGGCACTCGG CGGTCGAAAG GGGAGTTCAA GGAGACGGGG   60
GCGACGCGGC TGAGGGCTTC TCGTCGGGGT CGGGGCTGCA GCCGTCATGC CGGGGATAGT  120
GGAGCTGCCC ACTCTAGAGG AGCTGAAAGT AGATGAGGTG AAAATTAGTT CTGCTGTGCT  180
TAAAGCTGCG GCCATCACT ATGGAGCTCA ATGTGATAAG CCCAACAAGG AATTTATGCT  240
CTGCCGCTGG GAANAGAAAG ATCCGAGCGG GTGCTTAGAG GAAGGCAAAC TGGTCAACAA  300
GTGTGCTTTG GACTTCTTTA GGCAGATAAA ACGTCACTGT GCAGAGCCTT TTACAGAATA  360
TTGGACTTGC ATTGATTATA CTGGCCAGCA GTTATTTCTG CACTGTGCGA AACAGCAGGC  420
AAAGTTTGAC NAGTGTGTGC TGGACAACT GGGCTGGGTG CGGCCTGACC TGGGAAAACT  480
GTCAAAGTGC ACCAAAGTGA AAACAGATCN ACCTTTACCG GANAATCCCT ATCACTCAAG  540
AACAAGAACG GATCCCAGCC CTGANATCNA AGGAAATCTG CANCCTGCCA CACATGGCA   599

```

<210> 13

30

<211> 597

<212> DNA

<213> Homo sapiens

35

<400> 13

40

45

50

55

```

ATATCCGGAG TAGACGGAGC CGCAGTAGAC GGATCCGCGG CTGCACCAAA CACTGCCCCCT   60
CGGAGCCTGG TAGTGGGCCA CAAGCCCCCA GTCCCAGAGG CGTGATTTTC TGGCATCCTT  120
AAATCTTGTG TCAAGGATTG GTTATAATAT AACCAGAAAC CATGACGGCG GCTGAGAACG  180
TATGCTACAC GTTAATTAAC GTGCCAATGG ATTCAGAACC ACCATCTGAA ATTAGCTTAA  240
AAAATGATCT AGAAAAAGGA GATGTAAAGT CAAAGACTGA AGCTTTGAAG AAAGTAATCA  300
TTATGATTCT GAATGGTGAA AAACCTCCTG GACTTCTGAT GACCATCATT CGTTTTGTGC  360
TACCTCTTCA GGATCACACT ATCAAGAAAT TACTTCTGGT ATTTTGGGAG ATTGTTCTTA  420
AAACAACCTC AGATGGGAGA CTTTACATG AGATGATCCT TGTATGTGAT GCATACAGAA  480
AGGATCTTCA ACATCCTAAT GAATTTATTC NAAGGATCTA CTCTTCGTTT TCTTTGCAAA  540
TTGAAANAAA CANAATTGCT AAAACCTTTA ATGCCANCTA TNCCTGCATT TTTGGGA   597

```

<210> 14

<211> 634

<212> DNA

<213> Homo sapiens

<400> 14

```

AGACTCTCAC CGCAGCGGCC AGGAACGCCA GCCGTTACAG CGTTCGGTCC TCCTTGGCTG   60
ACTCACCGCC CTCGCCCGCC CACCATGGAC GCCCCCAGGC AGGTGGTCAA CTTTGGGCCT  120
GGTCCCGCCA AGCTGCCGCA CTCAGTGTG TTAGAGATAC AAAAGGAATT ATTAGACTAC  180
AAAGGANTTG GCATTAGTGT TCTTGAAATG AGTCACAGGT CATCAGATTT TGCCAAGATT  240
ATTAACAATA CAGAGAATCT TGTGCGGGAA TTGCTAGCTG TTCCAGACAA CTATAAGGTG  300
ATTTTCTGCG AAGGAGGTGG GTGCGGCCAG TTCAGTGCTG TCCCCTTAAA CCTCATTGGC  360
TTGAAAGCAG GAANGTGTGC GGAATATGTG GTGACAGGAG CTTGCTCAGC TAAGGCCGCA  420
NAANAAGCCA AGAANTTTGG GACTATAAAT ATCGTTCACC CTAAACTTGG GAGTTATACA  480
AAAATTCCAG ATCCAAGCAC CTGGAACCTC AACCAGATG CCTCCTACGT GTATTATTGC  540
GCNAATGAAA CNGTGCATGG TGTGGANTCT GACTTTATAC CCGATGTCNA GGAACATAC  600
TGTTTTGTGA CATGTCCTCA AACTTCCCGT CCNA                                634

```

<210> 15

<211> 757

<212> DNA

<213> Homo sapiens

<400> 15

```

AGTCTGCGGT GGGCTANCGG ACGGTCCGGC TTCCGGCGGC CGTTTCTGTC TCTTGCTGGC   60
TGTCTCGCTG AATCGCGGCC GCCTTCTCAT CGCTCCTGGA AGGTCCCGAG CGCGACACCA  120
TGTGCGAACC CGGGGGCGGC GCGGCGAAG ACNGCTCGGC CGGATTGGAA GTGTCGGCCG  180
TGCANAATGT GCGCGACGTG TCGGTGCTGC ANAAGCACCT GCGCAAGCTG GTGCCGCTGC  240
TGCTGGAGGA CGCGGCGGAA GCGCCGGCCG CGCTGGAGGC GCGCTGGAG GAGAAGAGCG  300
CCCTGGAGCA GATGCGCAAG TTCCTTTCCG ACCCGCACGT CCACACGGTG CTGGTGGAGC  360
GCTCCACGCT CAAAGTGGAC GTCGGTGATG AAGGAGAAGA AGAAAAAGAA TTCATTTCCT  420
ATAACATCAA CNTAGACATT CACTATGGGG TTAAATCCAA TAGCTTGGCA TTCATTAAAC  480
GTACTCCCGT GATTGATGCA GATAAACCGG TGTCTTCTCA NCTCCGGGTC CTTACACTCA  540

```



5 GTGAANACTC NCCCTACNAA AACTTTGCAT TCTTTCATTA ACAATGCAGT GGCTCCTTTT 600  
 TTTAANTCCT ACATTAAAAA ATCTGGCAAG GCAAACAGGG ATGGTGATAA AATGGCTCCT 660  
 TCCNTTGAAA AAAAAATTGC CGAACTCNAA ATNGGACTCC TTCCCTTGCA NCAAAATTTT 720  
 TGAAATTCCG GAAAATCANC CTGCCCAATT CCTCCCC 757

10 <210> 16

<211> 300

15 <212> DNA

<213> Homo sapiens

<400> 16

20 ATCAATTCCT TATTTATATT TCATGTTGGA ATGCTTAAAT CGATAACCTT TGTATTTTGA 60  
 AGTGCCGCAC ATGGAAGGTG ATCTGCAAGA GCTGCATCAG TCAAACACCG GGGATAAAAT 120  
 CTGGATTTGG GTTCCGGCGT CAAGGTGAAG ATAATACCTA AAGAGGAACA CTGTAAAATG 180  
 CCAGAAGCAG GTGAANAGCA ACCACAAGTT TAAATGAAGA CAAGCTGAAA CAACGCAAGC 240  
 25 TGGTTTTATA TTAGATATTT GACTTAAACT ATCTCAATAA AGTTTTGCAG CTTTCACCAC 300

<210> 17

30

<211> 313

<212> DNA

<213> Homo sapiens

35

<400> 17

40 AAAGATGGCG GCGGGGAGG TAGGCAGAGC AGGACGCCGC TGCTGCCGCC GCCACCGCCG 60  
 CCTCCGCTCC AGTCGCCTCC GGTCCCTCAA ACTCACACCT CCCGGGAGGA GCTGTCTTGG 120  
 CGCCGGGTCC CGCGGGGAAA ATGGTGGAGC CAGGGCAAGA TTTACTGCTT GCTGCTTTGA 180  
 GTGAGAGTGG AATTAGTCCG AATGACTCTT TGATATTGAT GGTGGAGATG CANGGCTTGC 240  
 AACTCCAATG CCTACCCCGT CAGTTCAGCA NTCAGTGCCA CTTANTGCAT TANAACANG 300  
 45 TTTGGAGACC GAA 313

<210> 18

50

<211> 667

<212> DNA

55

<213> Homo sapiens

<400> 18

```

ACTGCCGGGC TCGGCGTGAG TCGCTGCGGG GCTGACGGGG TGGCAGTGCG GCGGGTTACG   60
GCCTGGTCAG ACCATAATGA CTTCAGCAAA TAAAGCAATC GAATTACAAC TACAAGTGAA  120
ACAAAATGCA GAAGAATTAC AAGACTTTAT GCGGGATTTA GAAAACTGGG AAAAAGACAT  180
TAAACAAAAG GATATGGAAC TAAGAAGACA GAATGGTGTT CCTGAAGAGA ATTTACCTCC  240
TATTCGAAAT GGAATTTTAA GGAAAAAGAA GAAAGGCAAA GCTAAAGAGT CTTCCCCAAA  300
ACCANAGAGG AAAACACNAA AACAGGATA AAATCTTATG ATTATGANGC ATGGGCAAAA  360
CTTGATGTGG ACCGTATCCT TGATGAGCTT GACAAAGACG ATAGTACCCA TGAGTCTCTG  420
TCTCAAGAAT CAGAGTCGGA AGAAGATGGG ATTCATGTTG ATTCNCNAAA GGCTCTTGTT  480
TTAAAAGAAA AGGGCNATAA ATACTTCCAC AAGGAAAATA TGATGAAGCA ATTGACTGCT  540
ACACNAAAGG CNTGGATGCC GATCCATATN ATCCCGTGTT GCCAACGAAC ANAACNTCCG  600
CATATTTTAG ACTGAAAAAA TTTGCTGTTG CTGAATCTGA TTGTTATTAN CANTTGCCT  660
TGAATA                                           667

```

# Claims

1. A method for isolating a full-length cDNA clone, the method comprising:
  - (a) determining a nucleotide sequence from the 5'-region of a cDNA clone contained in a cDNA library;
  - (b) determining the presence or absence of an initiation codon in the nucleotide sequence determined in (a) using an initiation codon prediction program; and
  - (c) selecting clones recognized as containing the initiation codon in (b).
2. The method of claim 1, wherein the cDNA library is constructed by a method for preparing a full length-enriched cDNA library.
3. The method of claim 1, wherein a cDNA library is constructed by a method comprising a step of modifying Cap of mRNA.
4. A method for constructing a full length cDNA library, the method comprising:
  - (a) determining a nucleotide sequence from the 5'-region of a cDNA clone contained in a cDNA library;
  - (b) determining the presence or absence of an initiation codon in the nucleotide sequence determined in (a) using an initiation codon prediction program;
  - (c) selecting clones recognized as containing the initiation codon in (b); and
  - (d) combining the clones selected in (c).
5. The method of claim 4, wherein the cDNA library is prepared by a method for constructing a full length-enriched cDNA library.
6. The method of claim 4, wherein the cDNA library is constructed by a method comprising a step of modifying Cap of mRNA.
7. A cDNA library obtainable by the method of claim 4.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/04772

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> Int.Cl <sup>6</sup> C12N15/10  According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) Int.Cl <sup>6</sup> C12N15/10  Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) BIOSIS (DIALOG), WPI/L (DIALOG)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO, 94/08001, A1 (The Kanagawa Academy of Science), 14 April, 1994 (14. 04. 94) & JP, 6-153953, A & EP, 625572, A1	1-7
A	WO, 96/34981, A2 (GENSET), 7 November, 1996 (07. 11. 96) & EP, 824598, A2 & FR, 2733762, A1 & FR, 2733765, A1	1-7
A	Maruyama, K., et al., "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribo-nucleotides", Gene, Vol. 138 (1994), p.171-174	1-7
A	Kato, S., et al., "Construction of a human full-length cDNA bank", Gene, Vol. 150 (1994), p.243-250	1-7
A	Carnincle, P., et al., "High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper", (Genomics, Vol. 37 (1996), p.327-336	1-7
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" documents which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "A" document member of the same patent family		
Date of the actual completion of the international search 13 January, 1999 (13. 01. 99)		Date of mailing of the international search report 26 January, 1999 (26. 01. 99)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

Form PCT/ISA/210 (second sheet) (July 1992)

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/04772

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Edery, I., et al., "An Efficient Strategy To Isolate Full-Length cDNAs Based on an mRNA Cap Retention Procedure (CAPture)", Molecular and Cellular Biology, Vol. 15 (1995), p.3363-3371	1-7
A	Solovyev, V., et al., "Predicting internal exons by oligo-nucleotide composition and discriminant analysis of spliceable open reading frames", Nucleic Acids Research, Vol. 22 (1994), No. 24, p.5156-5163	1-7
A	Heindell, H.C., et al., "The Primary Sequence of Rabbit $\alpha$ -Globin mRNA", Cell, Vol. 15 (1978), p.43-54	1-7
A	Minoru Suzuki et al., "RT-PCR Process: Cloning of 5' end of mRNA by Oligocapping Procedure (in Japanese)", Protein, Nucleic Acid and Enzyme, Vol. 41, No. 5 (1996), p.603-607	1-7
A	Sumio Sugano et al., "Aiming at Full-length cDNA Library: Substitution of Capped Structure by Oligonucleotide (in Japanese)", Protein, Nucleic Acid and Enzyme, Vol. 38, No. 3 (1993), p.476-481	1-7
A	Carninci, P., et al., "High Efficiency Selection of Full-length cDNA by Improved Biotinylated Cap Trapper", DNA Research, Vol. 4, No. 1 (1997), p.61-66	1-7

Form PCT/ISA/210 (continuation of second sheet) (July 1992)